DOCUMENT RESUME

ED 433 720 FL 025 977

AUTHOR Tedick, Diane J.; Klee, Carol A.
TITLE Alternative Assessment in the Language Classroom.
INSTITUTION Center for Applied Linguistics, Washington, DC.
SPONS AGENCY Center for International Education (ED), Washington, DC.
PUB DATE 1998-12-00
NOTE 42p.; In its: Out of Class Learning Experiences and Students' Perceptions of Their Impact on English Conversation Skills; see FL 025 966.
CONTRACT P017A50050-95A
PUB TYPE Guides - Classroom - Teacher (052)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Check Lists; *College Instruction; Evaluation Methods; Graduate Students; Higher Education; Language Teachers; *Language Tests; Peer Evaluation; Portfolio Assessment; *Second Language Instruction; Self Evaluation (Individuals); *Student Evaluation; Teacher Education; *Teaching Assistants; Testing
IDENTIFIERS *Alternative Assessment

ABSTRACT

An instructional module designed to help prepare college-level teaching assistants (TAs) for their duties in second language instruction is presented. The module focuses on the use of alternative methods for assessing student learning. Introductory sections discuss the rationale for and definition of alternative assessment. Four specific types of alternative assessment are then defined and illustrated, with reference to recent research and practice. The four types include: creating speaking/listening, writing, and reading tasks that lend themselves to alternative assessment; using checklists and rubrics for assessing student performance on various language tasks; encouraging reflection through self-assessment and peer assessment; and use of portfolios of student work for assessment purposes. Specific examples and suggestions for classroom application for each assessment type are included. Contains 49 references. (MSE)

Alternative Assessment in the Language Classroom

Diane J. Tedick and Carol A. Klee
University of Minnesota

one of a series of modules for the

Professional Preparation of Teaching Assistants in Foreign Languages

Grace Stovall Burkart, Editor
Center for Applied Linguistics
Washington, DC

December 1998

Prepared under Grant Award No. PO17A50050-95A
Center for International Education
International Research and Studies Program
U.S. Department of Education

Alternative Assessment in the Language Classroom

Diane J. Tedick and Carol A. Klee
University of Minnesota

Table of Contents

Rationale for Alternative Assessment 2

Definition of Alternative Assessment 3

What are the challenges that come with this process? 3

The benefits that accompany the challenges 4

Types of Alternative Assessment 4

Creating tasks that lend themselves to alternative assessment 4

Using checklists and rubrics for assessing student performance on various language tasks 6

Encouraging reflection through self-assessment and peer assessment 17

Portfolio assessment 28

Concluding Thoughts: Advice to New Instructors 31

References 32

Suggested Additional Reading 35

Tables 29

Rationale for Alternative Assessment

As foreign language teachers we are expected by our students and our institutions to assess our students' progress in language acquisition. In many large language programs, course supervisors provide standardized midterm and final exams that focus on specific areas of the curriculum, e.g., listening comprehension, grammar, vocabulary, reading comprehension, speaking, and writing. Ideally, such tests should reflect the goals of the course as well as the way in which instruction in these areas is delivered and practiced in class.

In addition to traditional measures of language competence, alternative assessments have been developed in response to current interest in learner-centered pedagogy. Proponents of learner-centered pedagogy believe that teachers and learners should share power and that learners should have more control over their educational process (cf. Nunan, 1988). In this sense, the primary goal of learner-centered instruction is to increase students' participation in the learning process by assisting them in establishing learning and self-improvement goals, choosing effective learning methods and strategies, and becoming involved in evaluating their own work and that of their peers. Learner-centered instruction thus implies that teachers must dedicate some class time to activities not normally observed in traditional language classes, such as teaching learners how to learn a language, how to make use of available tools and resources, how to use language learning strategies, and how to reflect on their own learning. Language learners assume responsibilities traditionally taken on solely by the instructor, including the evaluation of their own learning, as well as the provision of feedback to their classmates.

Assessment procedures in any educational process should be congruent with teaching procedures. In other words, assessment practices should *align* with classroom objectives and instruction. If you are implementing a learner-centered approach in your classroom, you should consider using alternative assessment procedures as a further means of carrying out the approach. In keeping with the premises of learner-centered pedagogy, these assessment procedures are based on the idea that students can learn to evaluate their own learning and, in turn, learn from that process. They reflect the belief that learners should be involved in determining criteria for successful completion of communicative tasks and should have the opportunity to assess themselves and their peers. In addition, just as learner-centered pedagogy emphasizes both the learning *process* and the *product*, various forms of alternative assessment give learners opportunities to reflect not only on their linguistic development, but also on their learning processes (i.e., what helps them learn and what might help them learn better). Assessment thus becomes more formative rather than summative. Learners can provide one another with feedback on their performance, for example reflecting on how well they performed a communicative task through group processing (Johnson, Johnson, and Holubec, 1993). Finally, multiple measures (traditional as well as alternative forms) of student achievement are used to provide a more comprehensive picture of student performance.

We would like to emphasize that time spent on teaching students how to evaluate their own work through self-reflection and how to evaluate the work of their peers is not time lost for instruction. On the contrary, by understanding the traits of effective writers and speakers, students internalize the traits and become more effective communicators. As Baron (1991) states: "When students internalize a definition of what quality means and can learn to recognize it, they have developed a very valuable critical ability. They can talk with . . . their teacher about the quality of their work and take steps to acquire the knowledge and skills required to improve it" (p. 190).

Definition of Alternative Assessment

What, then, is alternative assessment? The search for alternatives to traditional types of assessment that primarily rely on pencil and paper tests (often requiring mere repetition of memorized material) has generated several innovative approaches to assessment having names like "performance assessment," "alternative assessment," and "authentic assessment" (Hart, 1994). These types of assessment are characterized by tasks that are worthwhile, significant, and meaningful, and form part of the curriculum. In the field of language education, this type of assessment provides information on what students can actually *do* with language and their reflection on that process. Hancock (1994) has defined it as "an ongoing process involving the student and teacher in making judgments about the students' progress in language using non-conventional strategies" (p. 1). It is congruent with a learner-centered, communicative approach to language teaching. Alternative assessments are not only designed and structured differently from traditional tests, but are also graded or scored differently. Student performance is evaluated on the basis of clearly defined performance indicators, criteria, or standards that emphasize students' strengths instead of highlighting their weaknesses.

What are the challenges that come with this process?

As with any change from an accustomed approach, the use of alternative assessments can create special challenges. First and foremost, you will need to read about and practice extensively with various forms of alternative assessments so that you become comfortable with them. At the same time, you will need to prepare your students for the use of alternative assessments. Learners who are used to traditional, teacher-centered classrooms may be reluctant to assume new roles and responsibilities. They may also be skeptical that peers can provide them with feedback that will enhance their learning. Be sure to explain the rationale for alternative assessment fully to learners. Indeed, you may find it beneficial to engage students in discussion about assessment in general and to elicit from them their thoughts on more traditional forms of assessment and their limitations with respect to assessing specifically what learners can *do* with language. Such a discussion may help students to understand the need for alternative assessment in the language classroom in conjunction with other forms. You will also need to provide students with guidance and instruction on how to reflect on their performance and evaluate it and how to evaluate their peers. We will give you some concrete suggestions on how to go about this below.

We would also like to emphasize the need to create a cooperative learning environment before attempting to use alternative assessments. Students must be in a supportive environment if they are expected to reflect thoughtfully on their learning processes. They must also feel comfortable with one another to provide constructive and honest feedback on their peers' work. Otherwise, they will provide perfunctory comments on other students' work to avoid hurt feelings.

For these reasons, it is important to introduce the use of alternative assessments gradually. Not only do instructors need to take time to become accustomed to these assessments, learners also need to understand how they will benefit from them and how they can use them effectively. Alternative assessments can easily be used alongside the more traditional means of assessment common to foreign language classrooms. A combination of alternative measures and more traditional forms of assessment makes it possible for the instructor to compare the results of the various approaches, leading to a more comprehensive picture of students' language performance than either alternative or traditional measures alone would provide. To allow students to become accustomed to them, we recommend that the instructor begin using checklists, scales, and rubrics (described in a subsequent section) to evaluate students' performance. This allows students to see the teacher modeling their use and gives them time to become accustomed to such assessments. Once students are familiar with the use of checklists, scales, and rubrics for evaluation, they can gradually begin to assess their own learning and provide feedback to their peers. Alternative assessments are generally designed to be an integral part or a natural culmination of a sequence of learning activities, but their use by both teachers and students requires careful preparation and should be implemented gradually.

The benefits that accompany the challenges

Changing the way we think about assessment simultaneously changes the way we think about teaching and the way students think about learning (Hart, 1994). This is perhaps one of the greatest benefits of alternative assessment—it focuses teachers' and students' attention on language *use*. Students become active participants in assessment activities that are designed to reveal what they can do with language rather than emphasizing their weaknesses. Teachers find alternative assessment techniques valuable in helping them to align instruction and assessment and in emphasizing for students communication for meaningful purposes.

Types of Alternative Assessment

Creating tasks that lend themselves to alternative assessment

Before introducing alternative assessment, it is essential to identify or design tasks that lend themselves to this type of assessment, i.e., those that provide students with the opportunity to demonstrate what they can actually do with language. This means that worksheets with fill-in-the-blanks exercises will not do the trick. Fortunately, many of the activities used in communicative classrooms lend themselves to this type of assessment:

Speaking/listening tasks — role-plays, interviews, group or individual presentations or demonstrations, debates, skits, information-gap activities

Writing tasks — journals, compositions, letters, e-mail correspondence or discussions, research reports

Reading tasks — skimming authentic texts for gist, scanning for specific information, comparing/contrasting articles or stories on the same topic written by different authors (or for different audiences)

As Baron (1991) states, "Many educators believe that performance-based assessments more closely represent the kinds of activities that we want our students to be able to undertake as members of society and that practicing for the assessment improves these valued skills and understandings" (p. 187). Certainly this is true in the case of language classrooms where students are learning to communicate in situations similar to those they will encounter in the "real world." Baron (1991) also points out that "There is a growing number of educators around the world who believe that there is little difference between an effective performance assessment task and an effective curriculum or learning task" (p. 191). This means that many of the activities that students engage in in a communicative classroom can be used as assessment tasks, although you should make sure to include a wide variety of task types that reflect real language use. "Implicit in this view of assessment is the need for the challenges we put before the student to better replicate the *interactive* challenges and standards of performance typically facing would-be language learners in the field as they 'do' their communication" (Wiggins, 1994, p. 71, emphasis in the original).

In designing communicative performance tasks for classroom use, it is important to keep in mind the notion of authenticity. Wiggins (1994, pp. 75-76) has proposed the following criteria to distinguish authentic from unauthentic forms of testing:

1. engaging and worthwhile problems or questions of importance . . .
2. faithful representation of the contexts facing workers in a field of study, or the real-life "tests" of adult life . . .
3. non-routine and multi-stage tasks, and real problems; recall or "plugging in" is insufficient or irrelevant . . .
4. tasks that require the student to produce a quality product and/or performance
5. transparent or de-mystified criteria or standards . . .
6. interaction between assessor and assessee . . .
7. provision for . . . concurrent feedback and the possibility of self-adjustment during the test . . .

Many tasks designed to develop communicative use of the second language fit these criteria.

In designing tasks, you should also consider authenticity in relation to the purpose of the task and its audience. Here we offer an example of a task that can be slightly altered to

become more authentic. Imagine that students are engaged in a unit on Costa Rica (or any other target country). As an assessment at the end of the unit, the teacher decides to have students create travel brochures in the target language (TL) to demonstrate their knowledge of what they have learned. Such a task asks that the students pretend to act as native speakers, which they clearly are not. Kramsch (1993) would argue that authenticity involves having students be who they are—learners of the TL. To revise the task somewhat with an eye toward greater authenticity, the teacher can have students create travel itineraries for a group of students who will be traveling to Costa Rica, the intent being to demonstrate their knowledge of what they have learned by communicating it to other students.

Another example for the same unit would involve having students at the beginning of the unit write letters in the TL to various travel agencies, tourist bureaus, and "Chamber of Commerce" equivalents to indicate that they (1) are students of Spanish, (2) are studying about Costa Rica, and (3) are interested in receiving travel information in Spanish. Such a task has a real purpose and a real audience. An added benefit is that it will also lead to additional authentic materials for classroom use.

A final example of an authentic task for this instructional setting is to have students write to Costa Rican students about Minnesota (i.e., their home state), given what they have learned about Costa Rica. A letter written for this task might include, for example, a comparison between Minnesota's Boundary Waters and Costa Rica's Tortuguero National Park in terms of their environmental restrictions. These suggestions highlight the importance of creating tasks that involve students in using language for real communicative purposes and for real audiences.

Using checklists and rubrics for assessing student performance on various language tasks

The use of checklists and rubrics is central to alternative assessment. Whereas a checklist simply provides an indication of whether a specific criterion, characteristic, or behavior is present, a rubric provides a measure of quality of performance on the basis of established criteria.

Checklists

Checklists are often used for observing performance or behavior in order to keep track of a student's progress or work over time. They can also be used to determine whether students have met established criteria on a task. Below is an example of a speaking task and a sample checklist (see Fig. 1) that might be used to check whether students meet the criteria needed to complete the task successfully.

Task description. For a unit on Hispanics in the United States, students are exploring issues related to Hispanics in Minnesota. They are instructed to make contact with a native Spanish speaker who has immigrated to Minnesota (teacher provides a list of resources for making contact). Students are to conduct a short interview with this individual and report back to the class. In an oral presentation, they are to (1) briefly describe the interviewee

(gender, age, place of birth, occupation, etc.); (2) explain what brought him/her to Minnesota; (3) describe at least one challenge the interviewee has faced or faces in Minnesota; (4) describe how this individual maintains a connection to his/her heritage; and (5) describe one item of interest that came out of the interview. Students are told that they will need to speak for a minimum of three minutes and that they are not to read to the class and can only refer to minimal notes while presenting. They are advised to rehearse, but not to memorize. A checklist for assessing students' completion of the task components might look like Figure 1. (See Figure 1 at the end of the text.)

- - - - - - - - - -

- - - - - - - - - -

Brown and Yule (1983) suggest a checklist-type scoring matrix for use with information-gap activities. The intention is to assess the speaker's communicative effectiveness. The first step is to select or create an information-gap task in which a speaker must describe or provide instructions to a listener, who follows the instructions or completes some task based on the description. For example, a speaker must explain to a listener how to assemble a mincer having five parts or components. The listener has the various parts of the mincer in front of him and is required to assemble the parts on the basis of the speaker's instructions. The speaker must be seated in such a way so that she cannot see what the listener is doing. The speaker begins by identifying the first part, then the second part and explains their relationship to one another, or how they fit together. She continues in this manner until all five parts are identified and their relationship with one another is described. While such tasks may not be considered "authentic" in the pure sense of the term, they do elicit the kinds of linguistic structures that students need to internalize during the process of language acquisition (Brown and Yule, 1983). A checklist for assessing the speaker's ability to communicate effectively is set up as in Figure 2. (See Figure 2 at the end of the text.)

- - - - - - - - - -

- - - - - - - - - -

The teacher listens to Speaker A's instructions and marks a check whenever she identifies a component and describes its relationship to another component. The same procedure is followed for Speaker B, C, etc. In the sample checklist in Figure 2, Speaker B was able to communicate all information effectively, whereas Speaker A's performance lacked some important details. In assessing communicative effectiveness, the teacher must be careful to listen to what the *speaker* says and not be influenced by what a listener does or does not do. That is, a listener may figure out a task and complete it without necessarily having explicit instructions from the speaker; conversely, the speaker may describe all of the required information and the listener may not follow the instructions correctly. Figure 2 may also be adapted to assess listening comprehension, in which case the teacher will pay attention to what a *listener* does on the basis of what a speaker says.

Checklists can be useful for classroom assessment because they are easy to construct and use, and they align closely with tasks. At the same time, they are limited in that they do not provide an assessment of the relative quality of a student's performance on a particular task.

Rubrics

In contrast to checklists, rubrics or scales provide an indication of *quality* of performance on a particular task. Rubrics have received much attention in recent years due to the increased emphasis on performance-based assessment. They are primarily used for language tasks that involve some kind of production on the part of the student, be it oral or written. Rubrics are created on the basis of four different scale types—holistic, analytic, primary trait, and multitrait—each of which was developed originally for large scale writing assessment. Scoring rubrics are often used with benchmarks, or samples that act as standards against which other samples are judged (Hart, 1994).

Holistic rubrics. When teachers use holistic scales or rubrics, they are responding to language performance (writing or speaking) as a *whole*. Each score on a holistic scale represents an overall impression; one integrated score is assigned to a performance. A well-known example of a holistic scale is the American Council on the Teaching of Foreign Languages (ACTFL) *Proficiency Guidelines* (1986). However, the ACTFL guidelines are not appropriate for classroom use, because they are intended for large-scale assessment of overall proficiency and are not designed necessarily to align with curricular objectives or classroom instruction.

The emphasis in holistic scoring is on what a student does well rather than what he or she has not done well (White, 1985). Holistic rubrics commonly have four or six points. Figure 3 shows a sample four-point holistic scale created for the purposes of assessing writing performance. (See Figure 3 at the end of the text.)

- - - - - - - - - -

- - - - - - - - - -

Holistic scoring is primarily used for large-scale assessment when a relatively quick yet consistent approach to scoring is necessary. It may be less useful for classroom purposes because it provides little information to students about their performance. Nevertheless, well-designed holistic scales provide for efficient scoring and may well be of value in classroom settings in addition to other forms of feedback.

Analytic rubrics. Analytic scales are divided into separate categories representing different aspects or dimensions of performance. Each dimension is scored separately, then dimension scores are added to determine an overall score. Common dimensions for writing performance include content, organization, vocabulary, grammar, and mechanics. On a scale

having these different categories, an essay would be evaluated by applying a different score to each category. This allows the teacher to weigh certain aspects more heavily than others. For example, content may have a total point range of 30 whereas mechanics may be attributed a total of 10 or 15 points.

One of the best known analytic rubrics used for writing assessment in the field of English as a second language (ESL) was developed by Hughey, Wormuth, Hartfiel, and Jacobs (1983). This rubric has five categories—content, organization, vocabulary, language use, and mechanics. Drawing heavily upon characteristics of the Hughey, et al. scale, Tedick and Klee developed an analytic rubric for use in scoring essays written for an immersion quarter for undergraduates studying Spanish (Klee, Tedick, and Cohen 1995). A recently revised version of the rubric appears in Figure 4. (See Figure 4 at the end of the text.)

Note that the scale in Figure 4 assigns different weights to different features. This allows an instructor to give more emphasis to content than to grammar or mechanics, for example. The option to weigh characteristics on the scale represents an advantage to analytic scoring. The decision to weigh certain criteria or not rests with the task, the purpose, and the level of the students. Figure 5 provides an example of an analytic scale that can be used for assessing speaking. This scale does not emphasize one feature over another, but certainly can be adapted to do so. (See Figure 5 at the end of the text.)

Analytic rubrics also have the advantage of providing more information to students about the strengths and weaknesses of various aspects of their language performance. One of the greatest criticisms of analytic scoring, however, is that the parts do not necessarily add up to the whole, or "the whole is greater than the sum of its parts." In other words, providing separate scores for different aspects of a student's writing or speaking performance may be considered artificial in that it does not give the teacher (or student) a good assessment of the "whole" of a performance.

Primary trait rubrics. The primary trait scoring method (Lloyd-Jones, 1977) involves predetermining the main criterion for successful performance on a task. The "primary trait" is defined by the teacher and varies depending upon the task. This approach thus involves narrowing the criteria for judging performance on a task to one main category or dimension. As an example, consider a task that requires that a student write a persuasive letter to an

editor of the school newspaper. The primary trait rubric might look something like Figure 6. (See Figure 6 at the end of the text. This kind of rubric has the advantage of allowing teachers (and students) to focus on one aspect or dimension of language performance. It is also a relatively quick and easy way to score writing or speaking performance—especially when a teacher wants to emphasize one specific aspect of that performance.

Multitrait rubrics. A multitrait approach to scoring language performance is similar to the primary trait approach but allows for rating performance on a number of dimensions (usually three or four) rather than emphasizing just one. Although similar to analytic rubrics in that several aspects are scored individually, multitrait rubrics are different in terms of the nature of the dimensions, or traits, that make up the scale. As explained above, an analytic scale comprises more traditional dimensions, such as content, organization, and grammar. A multitrait rubric, in contrast, involves dimensions that are more closely aligned with features of the task used to elicit language performance. For example, on an information-gap speaking task where students are asked to describe a picture in enough detail for a listener to choose it from a set of similar pictures, a multitrait rubric might be created that would include dimensions such as quality of description, fluency, and language control. (See Figure 7 at the end of the text).

In our multitrait example, the maximum total score is 12. Students are assigned a score of 1–4 for each of the three categories, and these are added to create a total score. The alignment of the scale with the task is perhaps the greatest strength of the multitrait rubric; at the same time this very alignment makes a multitrait rubric less transferable for use with other tasks. In other words, it is likely that each time a different task is used, a different rubric (or at least one or two dimensions of that rubric) will have to be developed.

Creating and using rubrics. While some scales or rubrics are created in such a way as to be generic in scope for use with any number of writing or speaking tasks, it is best to consider the task first and make sure that the rubric represents a good fit with the task and your instructional objectives. Just as a variety of task-types should be used in language classrooms, so should a variety of rubrics or scales be used for assessing performance on those tasks. Creating good rubrics that lend themselves well to consistent, accurate

assessments takes practice. It is a good idea to begin to collect samples of rubrics that you can refer to and borrow from in the process of developing your own.

Unlike traditional forms of assessment, which often involve more objective methods of scoring and grading, alternative assessments and their accompanying use of rubrics involve subjective judgments. This subjectivity makes it more challenging to establish reliability, or consistency, in scoring and grading. Although a thorough discussion of the notion of reliability as related to the use of rubrics used for performance assessment is beyond the scope of this module, a few pieces of advice can be offered. We recommend that you check your own reliability in some way. For example, as you grade students' written essays using a rubric, keep track of the scores you assign on a separate sheet of paper. A few days later, randomly select a number (e.g., five) of the essays and evaluate them again, being sure not to look at the original scores that you assigned. Then compare the two sets of scores to ensure that you assigned the same or nearly the same scores both times. If the two scores are quite different, you will need to examine the rubric carefully and re-evaluate the essays. This same procedure can be followed for checking your reliability in evaluating students' oral performance, as long as audio or video recordings of the performance are available. Also keep in mind that fatigue can affect an instructor's ability to score students' work consistently. It is a good idea, therefore, to limit the number of written essays or oral performances that you score at one sitting. The more practice you get with the rubrics and the more comfortable you become with the process, the more reliable your scoring will become. For a detailed discussion on reliability in scoring, see, for example, Cohen (1994).

Encouraging reflection through self-assessment and peer assessment

It has been suggested that good language learners are aware of language learning processes (e.g., Carrell, 1989; Devine, 1993; O'Malley and Chamot, 1990; Schmidt and Frota, 1986). They are aware of and able to reflect on their own and others' language learning strategies and progress as language learners. The ability to distance oneself from a situation and engage in deliberate thought about it defines reflection. While reflection has been recognized as important in all learning, it may be even more critical to language learning, because "the essence of second language education is embodied in its attempt to join individuals together so that they might communicate across linguistic and cultural boundaries" (Tedick, et al., 1993, p. 44). Our views of ourselves and our cultures and of those of others and their cultures are never uniform or static. As Kramsch (1991) explains,

> a large part of what we call culture is a social construct, the product of self and other[s'] perceptions."

Indeed, language use, or communication, is embedded always within culture, and therefore is largely dependent upon interlocutors' *perceptions* of meaning, which may or may not match the intended meaning. It is this very social, dynamic nature of language and culture that makes second languages different from and more special than other academic disciplines, and, hence, makes reflection so important.

Second language students should be provided with opportunities to engage in systematic reflection on a regular basis. Reflection requires commitment, time, and the will to be open, flexible, and sensitive. People need to begin with situations that they are

comfortable with and gradually build toward other more risk-taking ventures. Their reflection should be both culturally and linguistically based, as well as focused on self-as-learner and self-as-human-being. Students will not begin to engage in profound reflection on any of these levels overnight; the process needs to occur gradually and carefully, in an atmosphere where the students can ask questions freely (including those directed at a teacher's pedagogy!) and where risk-taking is encouraged (Tedick, 1992). One way to encourage reflection in students is to provide opportunities for them to assess their own language performance and that of others.

Self-assessment

The benefits of having students assess their own progress have been established in research on first-language literacy acquisition in young children (e.g., Brown, 1988; Glazer, 1992; Graves, 1983; Routman, 1991). It is believed that opportunities for self-assessment help students to become independent learners. In addition, a number of second-language studies have found that self-assessment leads to increased motivation in learners (Blanche and Merino, 1989). However, students do not learn to monitor or assess their learning on their own. Students must be taught strategies for self-monitoring and self-assessment. In the case of self-assessments, if time is not taken to instruct students in their use, their validity is questionable. Blanche and Merino (1989), in a review of sixteen studies that employed measures of self-assessment, found that among the factors that can threaten the validity of self-assessment was "the lack of common, valid criteria that both learners and instructors could use to make sound judgments" (p. 325) and learners' lack of training in how to perform the types of self-assessment that had been asked of them. Techniques for teaching students strategies for self-assessment are parallel to those used for teaching learning strategies. Detailed descriptions of such techniques can be found, for example, in O'Malley and Chamot's book on learning strategies (1990) and in Chamot's module on learning strategies in this series.

Self-assessment tools can be used to encourage students' reflection on topics they have studied, vocabulary they have learned, their study habits, and their sense of their overall strengths and weaknesses. Blanche and Merino (1989, pp. 338-340) suggest, for example, that students respond to the following kinds of questions:

1. In the past few lessons (days, weeks), we/I have studied/practiced/worked on: [Students are instructed to fill in a number of blanks with topics and areas (communicative functions, grammatical points, cultural aspects) relevant to their cases.]

2. In your estimation, how well can you deal with the topics you listed under question 1? [Students assess their performance or understanding by using a scale ranging from "not at all" to "thoroughly."]

3. On reflection, to what extent do you find the topics you listed under question 1 important in relation to your own needs? [Students respond by

using a scale ranging from "not important" to "extremely important."]

4. Questions 1–3 are repeated with an emphasis on vocabulary knowledge.
5. Summarizing the past few lessons, we/I feel that we/I have learned: [Students rate their learning on a scale ranging from "nothing at all" to "a lot."]

6. Looking back, I realize that I should change my study habits/learning approach/priorities in the following way:

7. Overall, I think my weaknesses are:

8. Overall, I think my strengths are: [This item was added to Blanche and Merino's list.]

9. In the next few lessons, I am interested in learning about:

Blanche and Merino suggest that students later share their self-assessments with a peer or in a small group, with instructions that they compare their impressions with other criteria such as test scores, teacher evaluations, and peers' opinions. This kind of practice is valuable in that it helps students to be aware of their learning; in addition, it not only informs the teacher about students' thoughts on their learning and progress, but also provides the teacher with feedback about course content and instruction.

Self-assessments can also be used to allow students to evaluate both language processes and products that are specific to the various modalities. Below we describe some techniques for getting at processes related to literacy development and cross-cultural awareness. We also suggest some ideas for involving students in the assessment of their performance.

Processes. Attempting to assess language learning processes represents a rather elusive endeavor. Nevertheless, it is possible to get a sense of students' processes through several self-assessment techniques. Here we describe three techniques: think-alouds, glossing during the writing process, and the use of journals for tapping into processes related to developing cross-cultural understanding.

"Think-alouds" or "verbal reports" can be produced by readers to provide a representation of the processes readers go through as they construct meaning from written text. In order to produce a think-aloud, a learner silently reads a portion of a text in the second language and says out loud (often in the first language) what she is thinking as she tries to construct meaning. She reads more, thinks out loud, and the process continues until the end of the text is reached. Think-alouds help learners and teachers get at *how* a learner goes about making sense of text. According to Glazer and Brown (1993, p. 89), think-alouds are valuable in that they (1) show some of a reader's in-process thinking; (2) encourage

thinking about text; (3) help teachers and learners to understand what confuses learners when reading; (4) can inform instructional decisions; and (5) have the potential to reveal to students who are working with peers how others think as they read.

The process approach to writing lends itself well to self-assessment, because students create multiple drafts, receive feedback (from teachers and/or peers), and revise their drafts based on the feedback. [For explanations of the process approach to writing as used in second-language contexts, see for example, Dvorak (1986), Hewins (1986), and Terry (1989).] One easy way to make students aware of their processes as they compose drafts of written text is to have them "gloss" one or two drafts. Glossing when applied to the writing process is similar to glossing unfamiliar vocabulary in a text by providing a definition of the unfamiliar terms in the margin. Glossing takes place after a student receives feedback on a draft. After a student composes a first draft, he or she receives feedback (usually in writing) from an instructor or peer about how to revise the draft to improve it. In the next draft, the student "glosses" his or her revisions by describing in the margins the changes made. A gloss might say, "expanded and developed this part more," or "provided a transition," or "rephrased this sentence." For beginning levels, glossing should occur in the native language. For more advanced learners, glossing might occur in the target language after students are taught the vocabulary and structures needed. Glossing not only helps the learner to focus on specific areas needing improvement, but also helps the instructor or peer reviewer to see exactly how the learner's writing develops from one draft to the next. Such a technique makes it easier for an instructor to assign a grade or award points for the effort that went into the writing process.

Winer and Steffensen (1992) provide a stimulating account of the benefits of peer dialogue journals for promoting cross-cultural awareness among beginning teachers. Language students can also keep personal or dialogue journals with peers to track their processes in gaining cultural understanding. Keeping a journal is important not only because the journal serves as a record, allowing one to see one's growth over time, but also because it involves the medium of writing. Reflection involves thinking, and thoughts change when they are put in writing. Writing provides a different form of reflection than speaking or thinking aloud offers.

Products. In addition to engaging in self-assessment of language and culture learning processes, students can also be asked to assess their own performance on language tasks and their cultural understanding or learning. Below we describe student-teacher contracts, goal setting, and having students rate their own language performance and cultural understanding using rubrics.

One way to begin the process of introducing students to self-assessment is to create student-teacher contracts. Contracts are written agreements between students and instructors, which commonly involve determining the number and type of assignments that are required for particular grades. For example, a student may agree to work toward the grade of "B" by completing a specific number of assignments at a level of quality described by the instructor.

Contracts can serve as a good way of helping students to begin to consider establishing goals for themselves as language learners.

Goal setting is an important part of self-assessment and learner-centered instruction. It is important for students to consider areas they wish to work on and to assess their progress in achieving a particular goal. Carolyn Tischer, a high school teacher of German in Osseo, Minnesota, has her students develop goals at the beginning of each grading period (personal communication, November, 1993). She learned early on that students have a tendency to create lofty long-range goals, such as "to speak German." In order to help students develop realistic, short-term, attainable goals, Tischer has used a framework which was developed by Lori Adam, a business teacher in the Osseo district, who used the SMART acronym originally created by Conlow (1991). (See Figure 8 at the end of the text.) Goal-setting, in addition to other forms of self-assessment, is also an important part of portfolio assessment, which we describe in more detail below.

Students can also take part in assessment by evaluating their own performance (and that of their peers) on the basis of checklists and rubrics that are developed. Earlier in this module we described the various types of checklists and rubrics that can be created by language teachers for assessing student performance on communicative, authentic tasks. Students can be taught how to rate their own performance by using such rubrics and checklists. In order to rate their own speaking performance, students would need to audio-tape or video-tape their performance and evaluate it using a rubric or checklist. Writing can easily be evaluated with rubrics.

We cannot emphasize too strongly the need to spend time with students to prepare them for self-assessment activities. Before asking students to rate their own or their peers' performance, you need to be sure that they understand the criteria and how to apply them. The more an instructor models and discusses the process, the more students will benefit from participating in the evaluation of their work.

In addition to participating in the assessment of their language performance, students need to be involved in assessing their cultural understanding and knowledge. If we consider that a major purpose of language education is to provide students with the knowledge and abilities to communicate across linguistic and cultural boundaries (Tedick, et al., 1993), we recognize the central role that culture needs to play in language education. Kramsch (1993) emphasizes that if we view language as social practice, we must see culture as the core of language instruction. It follows that if we agree that culture needs to be at the core of language instruction, we must also devise ways of assessing students' cultural knowledge and understanding. Wiggins (1989) and others have argued quite convincingly that if we value

something, we must assess it, for to neglect a concept in assessment is to communicate to students that the concept isn't important.

Kramsch (1993) has suggested that students need to learn about the multiplicity of perspectives that define cultural constructs. She argues that instead of having students simply state their interpretation of a cultural construct, they should be engaged in tasks that require them to *reflect an understanding* of a construct. For example, students have been learning about the educational system in Germany. Their task is to create a videotape about the U.S. educational system for a group of German students who will be on an exchange in the United States the following year. They are instructed to create a description of the U.S. system that reflects their understanding of what they have learned about the German system. In this way, teachers are able to tap into deeper levels of cross-cultural understanding.

We offer an example here of a performance task created for college-level students of French that includes a reflection of students' understanding of the French concept of "home." The task and assessments described below are intended to be interpreted as both teacher and student assessments. In other words, the tasks are designed in such a way as to allow for teacher assessment and students' self-assessment. This description incorporates many of the techniques and ideas we have discussed up to this point.

Suzanne Cook, a French instructor at the U.S. Air Force Academy, created this assessment for a course at the University of Minnesota (Cook, 1994). The performance task of this summative assessment is integrative in that it combines reading comprehension, writing, and cultural understanding. Before reading a text in French, students are instructed to reflect on their background knowledge of "the French and their homes" by responding to the following questions in English. They are assured that there are no right or wrong answers.

1. Describe the image you have of a French home. What is the image based on (TV, magazines, textbooks, visit to France—where in France, etc.)? In other words, reflect on what you believe has led you to form this image.

2. Would you characterize the French as hospitable to visitors in their home or not? Support your answer.

3. How would you describe Americans in terms of their hospitality? Feel free to use your own experience here: how does your family deal with guests in your home?

By beginning the assessment in this way, Cook communicates to students the value of using pre-reading strategies such as activating prior knowledge. She also gathers critical information that may help her understand a student's performance on the assessment. Next, students are instructed to read an excerpt from the book *Evidences Invisibles* (Carroll, 1987).

They are prompted with the following:

> The following excerpt comes from the book *Evidences Invisibles*, by Raymonde Carroll, a French anthropologist who is married to an American anthropologist and who has lived in the U.S. for some 20 years. She studied the common misunderstandings between French and American people, misunderstandings which are usually due to different assumptions about how one should live and which are not explicitly considered when individuals are interacting. The following passage reveals some of the fundamental assumptions which, according to Carroll, the French generally have about the home. Read the text carefully for understanding and with an eye for differences from your own concept of "home."

For assessing basic comprehension, Cook asks students to respond in English to some literal-level questions about the text. She also asks that they reflect in writing "on the author of this text and the implication this might have for the information she presents, in particular on how representative it might be of the whole population of France." By asking students to consider this inferential question, Cook attempts to tap students' understanding that the author's interpretation is directly related to her individual view of the world, based on her status and educational level, and that the information presented may not represent all French people. She assesses students' responses to this question with a checklist. (See Figure 9 at the end of the text.)

- - - - - - - - - -

- - - - - - - - - -

The basic comprehension questions and critical thinking/inferential question are followed by this performance task:

> Imagine you just received the following post card from a friend who recently arrived in Lyon to spend the summer with a French family. This friend is having some difficulty understanding the ways of his/her host family. With what you've learned from the reading passage, write a response to your friend in French to help him/her adjust. What should s/he do differently? Include information from the text (at least 3 main ideas), in your own words, and relate it to your friend's knowledge of the way Americans do things.

The postcard text is presented to the students in French, but its English translation is provided in Figure 10. (See Figure 10 at the end of the text.

- - - - - - - - - -

- - - - - - - - - -

The writing assignment represents an integrative task, where students are asked to link prior knowledge (of American homes and how Americans treat visitors in their homes) to new knowledge gained from the reading passage. A multitrait rubric having three categories (see Figure 11 at the end of the text) is used to assess the students' writing performance. Total scores may range from 3 to 12.

Peer assessment

One of the ways in which students internalize the characteristics of quality work is by evaluating the work of their peers. However, if they are to offer helpful feedback, students must have a clear understanding of what they are to look for in their peers' work. For example, when they read a peer's essay or listen to a presentation, should they focus only on grammatical accuracy? content? organization? or something else? The instructor must explain expectations clearly to them before they begin. If students are asked to give one another feedback on their essays, one way to make sure they understand what they are to evaluate is to provide students with a sample composition on an overhead and, as a group, determine what should be assessed (i.e., how does one define good writing), carry out the assessment, and then determine how to convey clearly to the fictitious student how he or she could improve the essay.

Students also benefit from the use of rubrics or checklists to guide their assessments; these rubrics can be provided by the instructor, or once the students have more experience, they can develop them themselves. Figure 12 (at the end of the text) is an example of a Peer Editing Checklist, which was developed by Susana Blanco-Iglesias, Joaquina Broner, Marisa Geisler, and Begoña Miguel-Pérez, and is used in second-year Spanish classes at the University of Minnesota-Twin Cities. Notice that the checklist requires that the peer evaluator comment primarily on the content and organization of the essay and helps the evaluator focus on aspects of good organization, e.g., "Are there concrete examples to support each of the ideas discussed?"

Before we introduced the Peer Editing Checklist, students complained that they did not receive feedback that was helpful from their peers and considered peer editing to be a waste of time. Now students write a first draft of their essay at home and turn it in the next day to a group of three peers. The peers then read the essay at home and fill out the checklist and return it to the student in the next class. The instructor provides fifteen to twenty minutes of class time for the peer editing groups to meet and discuss their suggestions

for each other's essays. Each student returns home with three peer reviews of his or her essay and then writes a second draft which is turned in to the instructor. The instructor provides feedback on content, organization, and grammatical accuracy. The student then writes the third and final draft but is evaluated on the entire process, including his/her evaluations of peers' essays. We believe that it is important to include peer evaluation in a student's final grade to insure that they take the task seriously.

In addition to peer assessment of writing, students can also evaluate their peers' oral presentations, role plays, skits, or debates. Again, it is important that students receive guidance on *what* to evaluate. The use of rubrics or checklists, as described earlier, helps students focus on the aspects that they should assess.

One final caveat: For peer evaluation to work effectively, the learning environment in the classroom must be supportive. Students must feel comfortable and trust one another to provide honest and constructive feedback. If you use process writing in your class and frequently use peer assessment, we recommend that you form groups of three to four students early in the semester and allow students to work within the same groups throughout the term. This will allow them to become more comfortable with each other and may lead to better feedback from peers.

Portfolio assessment

Definitions and characteristics of portfolios

Portfolios are purposeful, organized, systematic collections of student work that tell the story of a student's efforts, progress, and achievement in specific areas. This collection must include student participation in the selection of portfolio content, the guidelines for selection, and the criteria for judging merit (Hart, 1994; Tierney, et al., 1991). Portfolio assessment encompasses all that we have discussed thus far: an emphasis on a variety of tasks that elicit spontaneous as well as planned language performance for a variety of purposes and audiences, the use of rubrics to assess performance, and a strong emphasis on self-reflection and assessment (including goal setting), and peer assessment. Portfolio assessment lends itself well to meeting a variety of pedagogical objectives that are important to second language acquisition. Entire books have been written about portfolio assessment; it is an approach that requires a great deal of planning and collaboration. Here we offer a brief description of portfolio assessment, but we encourage readers to refer to the articles and books in our reference list for more detailed information before attempting this approach to assessment.

The following list of characteristics and functions of portfolios is adapted from a variety of sources, including: Evansville-Vanderburgh School Corp. (1993); Glazer and Brown (1993); Hart (1994); Hasty-Bambenek, Nielsen, and Tedick (1995); Tierney, et al. (1991). Portfolios:

- represent an emphasis on language use and cultural understanding;

- represent a collaborative approach to assessment;
- represent a student's range of performance in reading, writing, speaking, and listening as well as cultural understanding;
- emphasize what students can do rather than what they cannot do;
- represent a student's progress over time;
- engage students in establishing ongoing learning goals and assessing their progress towards those goals;
- measure each student's achievement while allowing for individual differences between students in a class;
- address improvement, effort, and achievement;
- allow for assessment of process and product; and
- link teaching and assessment to learning.

Fundamental to portfolio assessment is an emphasis on assessing students' progress, processes, and performance over time. It is important to distinguish between two basic types of portfolios. The first type may be referred to as an ongoing classroom or *process* portfolio and the second as a final *product* portfolio. Their distinction lies in the purpose that each serves. An ongoing classroom or *process* portfolio serves the purpose of classroom-level assessment on the part of both the instructor and the student. It is more reflective of *formative* assessment, although it may be assigned a grade at the end of a period of time, be it a semester or academic year. It may also include summative types of assignments that were awarded grades. In contrast, a final *product* portfolio, being more *summative* in nature, is intended for a major evaluation of some sort and is often accompanied by an oral presentation of the portfolio. For example, it may be used as a evaluation tool for graduation from a program or for the purpose of seeking employment.

To highlight the differences between these two types of portfolios, we offer the following examples. Rochelle Nielsen and Timothy Hasty-Bambenek teach high school Spanish and have their students keep ongoing portfolios during every academic year, which they carry with them to subsequent levels (Hasty-Bambenek, et al., 1995). At the beginning of the academic year, the portfolio assessment process is described to first year students at length, and they are asked to select three-ring binders (with pockets) for developing their portfolios. Students establish language/culture learning goals for the year and are asked to reflect on how the portfolio might help them achieve those goals. The portfolios are divided into a variety of sections that allow students to organize their assignments and reflections on those assignments. Students collect pieces or projects required by the teacher as well as work of their own choosing.

Hasty-Bambenek and Nielsen believe that the portfolios should be accompanied by individual conferences with students at various points throughout the academic year. These conferences allow students to discuss their goals, work, and progress toward the goals during the year. At the end of the year, students include a description of their final review process, which involves selection of significant pieces and reflections about why they believe they were successful (or not successful) with a particular piece. Nielsen (personal communication,

March, 1996) explains that such reflection helps students to focus on the relationship between effort and achievement. This emphasis on reflection is key, as it renders the portfolio more than a collection of work over time.

A good example of a final product portfolio is provided by Jeannette Bragger (1994), who developed with colleagues a series of procedures for assessing student outcomes in the French major at The Pennsylvania State University. In this context, the major portfolio is intended to represent a synthesis of students' learning during their studies in French. Students document both their progress in the use of French and their knowledge of Francophone literatures and cultures as well as their progress in analytical and critical thinking skills, problem-solving and synthesizing abilities, and research skills. After the student submits the final portfolio to his/her advisor, a meeting is scheduled for the student to present the portfolio orally to his/her three-member committee. The committee then judges the portfolio with a grade of "pass" or "pass with distinction."

At the postsecondary level, portfolios are most useful to students and instructors when they reflect extended periods of time, such as an entire academic year or two or more academic years. This suggests the need for a departmental commitment to this alternative form of assessment. Such commitment is apparent in Bragger's (1994) description of the major portfolio at Penn State and Fraser's (1995) description of Indiana University's approach to portfolio assessment in German.

Representing student progress on all modalities

Portfolios have the potential to represent student performance and progress on all language modalities. They can also be used, however, to reflect students' work on just one modality. They are most commonly linked with the writing process and have been used with great success in representing students' growth over time with literacy skills in elementary settings (e.g., Glazer and Brown, 1993).

Assessing portfolios

There are a variety of ways to assess portfolios, and the process can involve teacher assessment, students' personal assessment, collaborative assessment between the teacher and the student, peer assessment, or a combination thereof. Educators conducting a pilot project of portfolio assessment in foreign language classrooms in Indiana (Evansville-Vanderburgh School Corp., 1993) offer a variety of excellent guidelines for making decisions about the assessment process.

The evaluative process should include ongoing (formative) assessments of students' work as well as overall (summative) assessments. This overall assessment should require the students to select representative samples of their work attached to explanations as to how these selections best represent their progress. Students may also be asked to respond to questions that aren't necessarily tied to specific pieces in the portfolio, but instead reflect a general overall understanding. For example, they may be asked to explain what they have learned about the target culture during a particular time frame or to describe their own

contribution to their learning of the target language (Evansville-Vanderburgh School Corp., 1993).

Determining how to go about assessing portfolios in a systematic way is a process that involves reflection, much discussion and negotiation with students and colleagues, and risk-taking. The more the collaboration, the better the process, and, most certainly, the outcome.

Concluding Thoughts: Advice to New Instructors

First and foremost, we would like to emphasize once again that assessment must be congruent with both the content and goals of the curriculum and the processes of instruction. If you want to use alternative assessment effectively, your classroom should be communicative and learner centered.

Assuming that this is the case, one of the major challenges in the implementation of alternative assessment is time constraints. Particularly in large institutions with standardized exams, instructors feel that they must cover the material that will appear on the exams and do not have time to teach learners how to learn, how to use strategies, and how to assess themselves and their peers. If you are in this situation, we suggest that you speak with your course supervisor to determine if the amount of material covered can be reduced to allow sufficient time for attention to the learning process. However, if it is not possible to change the amount of material covered in the curriculum, you can still introduce some forms of alternative assessment in your classroom, especially if you develop ways of integrating them into the existing sequences of practice activities. (In this module see, for example, the writing assignments described on pp. 5-6, the oral activities on pp. 6-7, or the integrative reading and writing lesson on pp. 22-25.)

Finally, no matter how much flexibility you have in introducing alternative assessments, it is absolutely essential that you take the time to teach students how to use them. Introduce alternative assessment gradually and always in conjunction with more traditional forms of assessment. Students may not be accustomed to taking on responsibility for assessment and will need to adjust to this new role. As you become familiar with the process of creating authentic performance tasks for assessing language performance and of developing and using checklists and rubrics, and as your students grow accustomed to and find they benefit from evaluating themselves and their peers, you can expand the amount of alternative assessment used in your classroom.

References

American Council on the Teaching of Foreign Languages. (1986). *ACTFL Proficiency Guidelines*. Hastings-on-Hudson, NY: American Council on the Teaching of Foreign Languages.

Baron, J. B. (1991). *SEA Usage of Alternative Assessment: The Connecticut Experience.* (ERIC Document Reproduction Service No. ED 349 816).

Blanche, and B. J. Merino. (1989). Self-Assessment of Foreign Language Skills: Implications for Teachers and Researchers. *Language Learning*, 39(3): 313-340.

Bragger, J. (1994, October). *The Pennsylvania State University Department of French Assessment of Student Outcomes in the French Major*. Unpublished manuscript. University Park, PA: The Pennsylvania State University.

Brown, C. S. (1988). Think Aloud: A Method for Teaching the Processes of Comprehension. *The Reading Instruction Journal*, 32(1): 13-17.

Brown, and G. Yule. (1983). *Teaching the Spoken Language*. New York: Cambridge University Press.

Blanco-Iglesias, S., J. Broner, M. Geisler, and B. Miguel-Pérez. (n.d.). *Peer Editing Checklist*. Unpublished manuscript. Minneapolis, MN: Dept. of Spanish and Portuguese, University of Minnesota.

Carrell, P. L. (1989). Metacognitive Awareness and Second Language Reading. *Modern Language Journal*, 73: 121-133.

Carroll, R. (1987). *Evidences Invisibles: Américains et Français au Quotidien*. Paris: Seuil.

Cohen, A. D. (1994). *Assessing Language Ability in the Classroom*. 2nd ed. Boston, MA: Heinle & Heinle.

Conlow, (1991). *Excellence in Management*. Minneapolis, MN: Crisp Publications, Inc.

Cook, S. (1994, Fall). *The Concept of Home: Assumptions Not Shared*. Unpublished manuscript. Minneapolis, MN: University of Minnesota.

Devine, J. (1993). The Role of Metacognition in Second Language Reading and Writing. In J. G. Carson and I. Leki (Eds.), *Reading in the Composition Classroom*. Boston, MA: Heinle & Heinle, pp. 105-127.

Dvorak, T. (1986). Writing in the Foreign Language. In B. Wing (Ed.), *Listening, Reading,*

Writing: Analysis and Application. Northeast Conference Reports. Middlebury, VT: Northeast Conference, pp. 145-167.

Evansville-Vanderburgh School Corp. (1993). *Portfolio Assessment in Foreign Language, Pilot Project.* (ERIC Document Reproduction Services No. ED 368 197).

Fraser, Catherine. (1995). Portfolio Assessment in the Foreign Language Classroom: What Works. In G K. Crouse (Ed.), *Report of Central States Conference on the Teaching of Foreign Languages: Broadening the Frontiers of Foreign Language Education.* Lincolnwood, IL: National Textbook Co., pp. 98-106.

Glazer, S. M. (1992). *Reading Comprehension: Self-monitoring Strategies to Develop Independent Readers.* New York: Scholastic Professional Books.

Glazer, S. M. and C. S. Brown. (1993). *Portfolios and Beyond: Collaborative Assessment in Reading and Writing.* Norwood, MA: Christopher-Gordon Publishers, Inc.

Graves, D. H. (1983). *Writing: Teachers and Children at Work.* Portsmouth, NH: Heinemann.

Hancock, C. R. (1994, July). *Alternative Assessment and Second Language Study: What and Why?* ERIC Digest. (ERIC Document Reproduction Service No. ED 376 695).

Hart, D. (1994). *Authentic Assessment: A Handbook for Educators.* Reading, MA: Addison-Wesley Publishing Company.

Hasty-Bambenek, T., R. Nielsen, and D. J.Tedick. (1995, November). *Portfolio Assessment: Challenges and Inspirations.* Paper presented at the meeting of the American Council on the Teaching of Foreign Languages, Anaheim, CA.

Hewins, C. (1986). Writing in a Foreign Language: Motivation and the Process Approach. *Foreign Language Annals,* 19(3): 219-223.

Hughey, J. B., D. R. Wormuth, V. F. Hartfiel, and H. L. Jacobs. (1983). *Teaching ESL Composition: Principles and Techniques.* Rowley, MA: Newbury House.

Johnson, D. W., R. T. Johnson, and E. Holubec. (1993). *Cooperation in the Classroom.* Edina, MN: Interaction Book Company.

Klee, C. A., D. J. Tedick, and A. D. Cohen. (1995, March). *Content-Based Instruction in Spanish, French, and German at the University of Minnesota: The Development and Evaluation of a New Program.* Paper presented at the American Association of Applied Linguistics, Long Beach, CA.

Kramsch, C. (1991, May). *Toward a Pedagogy of Cross-cultural Competence.* Paper presented at an invited colloquium entitled Interdisciplinary Perspectives on Culture Learning in the Second Language Curriculum, University of Minnesota, Minneapolis, MN.

Kramsch, C. (1993). *Context and Culture in Language Teaching.* New York: Oxford University Press.

Lloyd-Jones, R. (1977). Primary Trait Scoring. In C. R. Cooper and L. Odell (Eds.), *Evaluating Writing.* Urbana, IL: National Couuncil on the Teaching of English, pp. 33-66.

Nunan, David. (1988). *The learner-centered curriculum: A study in second language teaching.* New York: Cambridge University Press.

O'Malley, J. M. and A. U. Chamot. (1990). *Learning Strategies in Second Language Acquisition.* New York: Cambridge University Press.

Routman, R. (1991). *Invitations: Changing as Teachers and Learners K-12.* Portsmouth, NH: Heinemann.

Schmidt, R. W. and S. N. Frota. (1986). Developing Basic Conversational Ability in a Second Language: A Case Study of an Adult Learner of Portuguese. In R. R. Day (Ed.), *Talking to Learn.* Rowley, MA: Newbury House, pp. 237-326.

Shohamy, E. (1985). *A Practical Handbook in Language Testing for the Second Language Teacher.* Israel: Shoshana Goldberg.

Terry, R. M. (1989). Teaching and Evaluating Writing as a Communicative Skill. *Foreign Language Annals,* 22(1): 43-54.

Tedick, D. J. (1992). Reflections on Reflection. *Elsie Speaks,* 3(1): 4-6. [Newsletter of the Language Center at the University of Minnesota.]

Tedick, D. J., C. L. Walker, D. L. Lange, R. M. Paige, and H. L. Jorstad. (1993). Second Language Education in Tomorrow's Schools. In G. Gunterman (Ed.), *Developing Language Teachers for a Changing World.* ACTFL Foreign Language Education Series. Lincolnwood, IL: National Textbook Company, pp. 43-75.

Tierney, R. J., M. A. Carter, and L. E. Desai. (1991). *Portfolio Assessment in the Reading-Writing Classroom.* Norwood, MA: Christopher-Gordon Publishers, Inc.

Tischer, C. (n.d.). *SMART Goals.* Unpublished manuscript. Osseo, MN.

White, E. (1985). *Teaching and Assessing Writing*. San Francisco: Jossey-Bass.

Wiggins, G. (1989). A True Test: Toward More Authentic and Equitable Assessment. *Phi Delta Kappan, May 1989,* 703-704.

Wiggins, G. (1994). Toward More Authentic Assessment of Language Performance. In C. R. Hancock (Ed.), *Teaching, Testing, and Assessment: Making the Connection*. Northeast Conference Reports. Lincolnwood, IL: National Textbook Company, pp. 69-85.

Winer, L. and M. S. Steffensen. (1992). Cross-cultural Peer Dialogue Journals in ESOL Teacher Education. *TESOL Journal,* 1(3): 23–27.

Suggested Additional Reading

Alternative Assessment. (1995). Special issue of the *TESOL Journal*, 5(1).

This entire issue of *TESOL Journal* is devoted to the theme of alternative assessment and contains very practical articles about using portfolios, self-assessment techniques, and other innovative assessment practices in a variety of second language contexts. Although directed to English as a second language teachers, the issue should prove to be of interest to foreign language instructors as well.

Belanoff, P. and M. Dickson. (Eds.). (1991). *Portfolios: Process and Product*. Portsmouth, NH: Boynton/Cook Publishers, Inc., Heinemann.

This edited volume contains chapters by well-known scholars in the field of first language writing. It is divided into four sections: portfolios for proficiency testing, program assessment, classroom portfolios, and political issues. The chapters offer both theoretical insights and ideas for practical application.

Brindley, G. (1989). *Assessing Achievement in the Learner-Centred Curriculum*. Sydney, Australia: National Centre for English Language Teaching and Research, Macquarie University.

This book provides a detailed synopsis of a study commissioned by the Adult Migrant Education Program in Australia for examining ways of assessing the achievement of adult learners of English as a second language. The report summarizes the results of the study and provides in-depth descriptions of assessment procedures and rating scales that emphasize criterion-referenced methods.

Hancock, C. R. (Ed.). (1994). *Teaching, Testing, and Assessing: Making the Connection*. Northeast Conference Reports. Lincolnwood, IL: National Textbook Co.

This volume of the Northeast Conference Reports examines the relationship between teaching, testing, and assessment in the foreign language classroom. It includes chapters that focus on the assessment of specific language modalities (i.e., speaking, listening, reading, and writing) as well as those that take into account affective considerations and the assessment of culture through portfolios. The chapter by Wiggins provides an excellent rationale for the use of alternative assessments in foreign language teaching.

Gill, K. (Ed.). (1993). *Process and Portfolios in Writing Instruction*. Urbana, IL: National Council of Teachers of English.

In this collection of articles, English language arts teachers representing elementary, secondary, and college settings, offer ideas for implementing a process approach to the teaching of writing in conjunction with the use of portfolios for assessment. The contributors emphasize the value of student collaboration and student responsibility in the process of learning to become effective writers.

Smith, M. A. and M. Ylvisaker. (Eds.). (1993). *Teachers' Voices: Portfolios in the Classroom.* Berkeley, CA: National Writing Project.

In this collection, thirteen classroom teachers discuss the difficulties and benefits of using portfolios for writing assessment.

The Authors

Carol A. Klee is Chair of the Department of Spanish and Portuguese and Associate Professor of Hispanic Linguistics at the University of Minnesota, where she directed the Spanish and Portuguese language programs from 1985 to 1995. She is Past President of the American Association of University Supervisors, Coordinators, and Directors of Language Programs and has served on the Executive Council of the American Association of Teachers of Spanish and Portuguese. Her research interests include sociolinguistics, language contact, and second language acquisition. She has edited several collections of research, including *Sociolinguistics of the Spanish-Speaking World* (1991) and *Faces in a Crowd: The Individual Learner in Multisection Courses* (1994).

Diane J. Tedick is Associate Professor in Second Languages and Cultures Education in the Department of Curriculum and Instruction, College of Education and Human Development, at the University of Minnesota. She co-directs the preservice teacher preparation program and offers courses for inservice teachers and other graduate students on topics such as second language reading and writing, curriculum development, and assessment. She is the coordinator of the Teacher Development Project of the University's National Foreign Language Resource Center. Her work has appeared in *Modern Language Journal, Foreign Language Annals*, and *Teaching Education.* She has also co-authored chapters in a variety of books, including the ACTFL Foreign Language Education series and the Northeast Conference reports.

Figure 1 Checklist for Oral Presentation of Interview

Criteria	Yes	No
Describes interviewee (gender, age, place of birth, occupation).		
Explain's interviewee's immigration to Minnesota.		
Describes at least one challenge the interviewee faces.		
Describes how interviewee maintains connection to culture.		
Describes point of interest.		
Speaks for a minimum of 3 minutes.		
Evidence of rehearsal (not reading to class).		

Figure 2 Checklist for Information-Gap Exercise

Required Information	Speaker A	B	C	D....
component 1	√	√		
component 2	√	√		
relationship between 2 and 1		√		
component 3	√	√		
relationship between 3 and 2/1		√		
component 4	√	√		
relationship between 4 and 3/2/1	√	√		
component 5	√	√		
relationship between 5 and the rest		√		

Adapted from Brown and Yule (1983). Reprinted by permission of Cambridge University Press.

Figure 3 Holistic Scale for Assessing Writing

4	**Excellent**—Communicative; reflects awareness of sociolinguistic aspects; well-organized and coherent; contains a range of grammatical structures with minor errors that do not impede comprehension; good vocabulary range.
3	**Good**—Comprehensible; some awareness of sociolinguistic aspects; adequate organization and coherence; adequate use of grammatical structures with some major errors that do not impede comprehension; limited vocabulary range.
2	**Fair**—Somewhat comprehensible; little awareness of sociolinguistic aspects; some problems with organization and coherence; reflects basic use of grammatical structures with very limited range and major errors that at times impede comprehension; basic vocabulary used.
1	**Poor**—Barely comprehensible; no awareness of sociolinguistic aspects; lacks organization and coherence; basic use of grammatical structures with many minor and major errors that often impede comprehension; basic to poor vocabulary range.

Adapted from Cohen (1994), reprinted by permission of Heinle & Heinle Pubishers, and from Shohamy (1985), reprinted by permission of the author.

Figure 4 Analytic Writing Scale for the Spanish FLIP Program

Content — 30 Total points possible

Score	Range	Criteria	Comments
	30 – 27	**Excellent to Very Good**— ◦ addresses all aspects of the prompt ◦ provides good support for and development of all ideas with range of detail ◦ substantive	
	26 – 22	**Good to Average**— ◦ prompt adequately addressed ◦ ideas not fully developed or supported with detail, though main ideas are clear ◦ less substance	
	21 – 17	**Fair**— ◦ prompt may not be fully addressed (writer may appear to skirt aspects of prompt) ◦ ideas not supported well, main ideas lack detailed development ◦ little substance	
	16 – 13	**Poor**— ◦ doesn't adequately address prompt ◦ little to no support or development of ideas ◦ non-substantive	

Organization — 20 Total points possible

Score	Range	Criteria	Comments
	20 – 18	**Excellent to Very Good**— ◦ well-framed and organized (with clear introduction, conclusion) ◦ coherent ◦ succinct ◦ cohesive (excellent use of connective words)	
	17 – 14	**Good to Average**— ◦ adequate but loose organization with introduction and conclusion (though they may be limited or one of the two may be missing) ◦ somewhat coherent ◦ more wordy rather than succinct ◦ somewhat cohesive (good use of connective words)	
	13 – 10	**Fair**— ◦ lacks good organization (no evidence of introduction, conclusion) ◦ ideas may be disconnected, confused ◦ lacks coherence ◦ wordy and repetitive ◦ lacks consistent use of cohesive elements	
	9 – 7	**Poor**— ◦ confusing, disconnected organization ◦ lacks coherence, so much so that writing is difficult to follow ◦ lacks cohesion	

Figure 4 continued on next page

Figure 4 Analytic Writing Scale (cont.)

Language Use/Grammar/Morphology — 25 Total points possible

Score	Range	Criteria	Comments
	25 – 22	**Excellent to Very Good**— ◦ great variety of grammatical forms (e.g., range of indicative verb forms; use of subjunctive) ◦ complex sentence structure (e.g., compound sentences, embedded clauses) ◦ evidence of "Spanish-like" construction ◦ mastery of agreement (subj/verb; number/gender) ◦ very few errors (if any) overall with none that obscure meaning	
	21 – 18	**Good to Average**— ◦ some variety of grammatical forms (e.g., attempts, though not always accurate, at range of verb forms, use of subjunctive) ◦ attempts, though not always accurate, at complex sentence structure (e.g., compound sentences, embedded clauses) ◦ little evidence of "Spanish-like" construction, though without clear translations from English ◦ occasional errors with agreement ◦ some errors (minor) that don't obscure meaning	
	17 – 11	**Fair**— ◦ less variety of grammatical forms (e.g., little range of verb forms; inaccurate, if any, attempts at subjunctive) ◦ simplistic sentence structure ◦ evidence of "English-like" construction (e.g., some direct translation of phrases) ◦ consistent errors (e.g., with agreement), but few of which may obscure meaning	
	10 – 5	**Poor**— ◦ very little variety of grammatical forms ◦ simplistic sentence structure that contains consistent errors, especially with basic aspects such as agreement ◦ evidence of translation from English ◦ frequent and consistent errors that may obscure meaning	

Figure 4 continued on next page

Figure 4 Analytic Writing Scale (cont.)

Vocabulary/Word Usage — 20 Total points possible

Score	Range	Criteria	Comments
	20 – 18	**Excellent to Very Good**— ◦ sophisticated, academic range ◦ extensive variety of words ◦ effective and appropriate word/idiom choice and usage ◦ appropriate register	
	17 – 14	**Good to Average**— ◦ good, but not extensive (less academic) range or variety ◦ occasional errors of word/idiom choice or usage (some evidence of invention of "false" cognates), but very few or none that obscure meaning ◦ appropriate register	
	13 – 10	**Fair**— ◦ limited and "non-academic" range (frequent repetition of words) ◦ more consistent errors with word/idiom choice or usage (frequent evidence of translation; invention of "false" cognates) that may (though seldom) obscure meaning ◦ some evidence of inappropriate register	
	9 – 7	**Poor**— ◦ very limited range of words ◦ consistent and frequent errors with word/idiom choice or usage (ample evidence of translation) ◦ meaning frequently obscured ◦ evidence of inappropriate register	

Mechanics — 5 Total points possible

Score	Range	Criteria	Comments
	5	**Excellent to Very Good**— ◦ demonstrates mastery of conventions ◦ few errors in spelling, punctuation, capitalization, and use of accents	
	4	**Good to Average**— ◦ occasional errors in spelling, punctuation, capitalization, and use of accents, but meaning is not obscured	
	3	**Fair**— ◦ frequent errors in spelling, punctuation, capitalization, and use of accents that at times confuses or obscures meaning	
	2	**Poor**— ◦ no mastery of conventions ◦ dominated by errors in spelling, punctuation, capitalization, and use of accents	

Total Score: Comments:

University of Minnesota. Revised July 1996.
Adapted from Hughey, et al. (1983, p. 140). Reprinted by permission of the authors.

Figure 5 Analytic Scale for Assessing Speaking

Pronunciation	
4	**Excellent**—No consistent or conspicuous mispronunciation; approaches native-like pronunciation with good intonation and juncture.
3	**Good**—Some identifiable deviations in pronunciation, but with no phonemic errors. Non-native accent evident with occasional mispronunciations that do not interfere with understanding.
2	**Fair**—Identifiable deviations in pronunciation with some phonemic errors. Non-native accent requires careful listening, and mispronunciations lead to occasional misunderstanding.
1	**Poor**—Frequent pronunciation errors with a heavy non-native accent. Many phonemic errors that make understanding difficult.
Fluency	
4	**Excellent**—Speech is effortless and smooth with speed that approaches that of a native speaker.
3	**Good**—Speech is mostly smooth but with some hesitation and unevenness caused primarily by rephrasing and groping for words.
2	**Fair**—Speech is slow and often hesitant and jerky. Sentences may be left uncompleted, but speaker is able to continue, however haltingly.
1	**Poor**—Speech is very slow and exceedingly halting, strained, and stumbling except for short or memorized expressions. Difficult for a listener to perceive continuity in utterances and speaker may not be able to continue.
Grammar/Language Use	
4	**Excellent**—Very strong command of grammatical structure and some evidence of difficult, complex patterns and idioms. Makes infrequent errors that do not impede comprehension.
3	**Good**—Good command of grammatical structures but with imperfect control of some patterns. Less evidence of complex patterns and idioms. Limited number of errors that are not serious and do not impede comprehension.
2	**Fair**—Fair control of most basic syntactic patterns. Speaker always conveys meaning in simple sentences. Some important grammatical patterns are uncontrolled and errors may occasionally impede comprehension.
1	**Poor**—Any accuracy is limited to set or memorized expressions; limited control of even basic syntactic patterns. Frequent errors impede comprehension.

Figure 5 continued on next page

Figure 5 Analytic Scale for Assessing Speaking (cont.)

Vocabulary	
4	**Excellent**—Very good range of vocabulary with evidence of sophistication and native-like expression. Strong command of idiomatic expressions. Infrequent use of circumlocution because particular words are rarely lacking.
3	**Good**—Good range of vocabulary with limited evidence of sophistication. Some expressions distinctly non-native but always comprehensible. Limited evidence of idiomatic expressions. Speaker is comfortable with circumlocution when lacking a particular word.
2	**Fair**—Adequate range of vocabulary with no evidence of sophistication. Some distinctly non-native expressions or errors in word choice may impede comprehension. No evidence of idiomatic expressions. Speaker has difficulty with circumlocution when lacking a particular word.
1	**Poor**—Limited range of vocabulary. Lack of repertoire and frequent errors in word choice often impede comprehension. Speaker shows no attempt at circumlocution when lacking a particular word.

Total Score:

Adapted from Shohamy (1985, pp. 183-184). Reprinted by permission of the author.

Figure 6 Primary Trait Rubric

Primary Trait: Persuading an audience

0 — Fails to persuade the audience.

1 — Attempts to persuade but does not provide sufficient support.

2 — Presents a somewhat persuasive argument but without consistent development and support.

3 — Develops a persuasive argument that is well developed and supported.

Figure 7 Multitrait Rubric

	Quality of Description	Fluency	Language Control
4	High level of accuracy in description is reflected; high degree of detail included in description.	Smooth and fluid speech; few to no hesitations; no attempts to grope for words.	Excellent control of language features; a wide range of well chosen vocabulary; accuracy and variety of grammatical structures.
3	Good accuracy in description, though some detail might be lacking.	Speech is relatively smooth but is characterized by some hesitation and unevenness caused by rephrasing and/or groping for words.	Good language control; good range of relatively well chosen vocabulary; some errors in grammatical structures possibly caused by attempt to include a variety.
2	Description lacks some accuracy and some critical details are missing that make it difficult for the listener to complete the task.	Speech is frequently hesitant and jerky, with some sentences left uncompleted.	Adequate language control; vocabulary range is lacking; frequent grammatical errors that do not obscure meaning; little variety in structures.
1	Description is so lacking that the listener cannot complete the task.	Speech is slow and exceedingly hesitant and strained except for short or memorized phrases; difficult to perceive continuity in utterances.	Weak language control; basic vocabulary choice with some words clearly lacking; frequent grammatical errors, even in simple structures, that at times obscure meaning.

Total Score:

Adapted from Cohen (1994), reprinted by permission of Cambridge University Press, and from Shohamy (1985), reprinted by permission of the author.

Figure 8 Setting SMART Goals

GOAL(S)	State your goal(s) below. Examine your goal(s).
Specific	Is it focused? Explain.
Measurable	Can you tell if it is accomplished? Explain.
Attainable	Is it a realistic target, given the time frame? Explain.
Relevant	Is it a priority? Explain.
Trackable	Can the results be compared over time? Explain.
OUTCOME	Describe whether and how you achieved your goal(s).

[illegible] n.d.). Used by permission of the author.

Figure 9 Assessing Responses to an Inferential Question

Situates author as educated and/or (at least) middle class.	Yes ___ No ___
Demonstrates an understanding that information might not/does not represent all French people, or more generally that social variables affect the way people behave.	Yes ___ No ___

(Cook, 1994). Reprinted by permission of the author.

Figure 10 Information Input for a Writing Task

Dear _____,

I just arrived at the Fourniers' house, and I seem to have begun my stay with them on the wrong foot! The family prepared a dinner to celebrate my arrival and invited some friends. I decided to help Mrs. Fournier in the kitchen, but she insisted that I leave and stay out of the kitchen. Later, I greeted some guests at the door with Mr. Fournier and was happy to help by taking the woman's coat and putting it on the bed in Mr. and Mrs. Fournier's room. But when I came out of the room, Mrs. Fournier had a surprised look on her face and didn't seem very pleased. Later on, so as not to bother Mr. or Mrs. Fournier, I went into the kitchen and grabbed a beer out of the fridge. When I returned to the living room, Mr. and Mrs. Fournier seemed completely shocked. I truly cannot understand what I did to make them so angry.

Tell me what you think. Please write soon!

Michael/Michelle

(Cook, 1994). Reprinted by permission of the author.

Figure 11 Assessing Performance on a Writing Task

	Content	Cultural Sensitivity	Language Control
4	Writing reflects thorough comprehension of the reading passage; effectively addresses the topic (is convincing to a reader); mentions at least 3 main ideas from the reading passage as support; demonstrates integration of new and prior knowledge.	Ideas expressed in the writing about the target culture avoid making a judgment as to whether the target culture (e.g., France) or home culture (e.g., U.S.) is better or worse.	Excellent control of language features; a wide range of well-chosen vocabulary and appropriate register; accuracy and variety of grammatical structures; uses own words to convey ideas from the reading passage.
3	Writing reflects good comprehension of the reading passage; adequately addresses the topic; mentions at least two main ideas from the reading passage as support; demonstrates attempts at integration of new and prior knowledge.	Ideas expressed in the writing about the target culture generally avoid making judgment as to whether the target culture (e.g., France) or home culture (e.g., U.S.) is better or worse, though some language used might suggest judgment. Less, rather than more, judgmental.	Good language control; good range of relatively well-chosen vocabulary; appropriate register; some errors in grammatical structures possibly caused by attempt to include a variety; clear attempts to use own words to convey ideas from the reading passage.
2	Writing reflects some comprehension of the reading passage; fairly addresses the topic, though may miss some critical points; mentions at least one main idea from the reading passage as support; demonstrates attempts at integration of new and prior knowledge, but writing might reflect some misunderstanding	Ideas expressed in the writing about the target culture at times seem to reflect judgment as to whether the target culture (e.g., France) or home culture (e.g., U S.) is better or worse. More, rather than less, judgmental.	Adequate language control; vocabulary range is lacking; register may/may not be consistently appropriate. Frequent grammatical errors that do not obscure meaning; little variety in structures. Doesn't always attempt to use own words to convey ideas from reading passage (has "lifted" portions).
1	Writing does not consistently reflect comprehension of the reading passage; topic is not adequately addressed and critical points are missing; little to no support from reading passage; writing reflects some misunderstanding.	Ideas expressed in the writing about the target culture often reflect judgment as to whether the target culture (e.g., France) or home culture (e.g., U.S.) is better or worse. Very judgmental.	Weak language control; basic vocabulary choice with some words clearly lacking; frequent grammatical errors even in simple structures that at times obscure meaning. Inconsistent use of register. Consistently "lifts" large portions of reading passage rather than attempting to use own words.

(Cook, 1994). Reprinted by permission of the author.

Figure 12 Peer Editing Checklist

Department of Spanish and Portuguese
University of Minnesota

PEER EDITING CHECKLIST

CONTENT AND ORGANIZATION
Composition/Draft #___

Editor _____ Student's name _____

I. What is the **main point** of the composition?

Is it clearly stated in the introduction?

Are there irrelevant, redundant, ambiguous, or otherwise unclear ideas present? Where?

Are there concrete examples to support each of the ideas discussed? Where are more examples needed?

What changes need to be made in order to clarify and/or develop the main point better?

II. Are all three of the "Basic Components of a Composition" (introduction, support conclusion) present?

Which of these parts need(s) further development and in what way?

III Is there a title?

Are the paragraphs logically related and sequenced? If not, where?

Is there adequate use of *connectors* between sentences and also paragraphs?

Is each idea/sub-topic organized into a different paragraph?

IV. Are there passages in which spelling and/or grammar hinder the meaning? Where?

GENERAL COMMENTS:

Blanco-Iglesias, et al. (n.d.). Used by permission of the authors.